**重庆邮电大学本科毕业设计（论文）开题报告**

|  |  |  |  |
| --- | --- | --- | --- |
| 题 目 | 基于语音识别与分析的人体管理系统设计 | | |
| 学生姓名 | 朱孝虎 | 学 号 | 2017212965 |
| 指导教师 | 邓钦元 | 所在单位 | 自动化学院工业互联网学院 |
| 一、选题背景（综述本课题研究现状、选题目的及意义） | | | |
| **1.1历史背景**  话语作为人们交流表达最直接的方式，其中蕴藏了大量的信息，能体现出你实时情感，当前任务情况以及未来的事等等，如果我们能把这些数据收集起来并做好相应的分析，既能增强对自己的了解，也能做好更好的选择。同时这个项目也算是分析记录人体数据做出的尝试。  当前社会上流通的只有功能单一的录音机，只能单机的进行语音的收集和简单的存储，并没有对这些数据进行合理的存储和分析并得到有价值的东西。  所以这个系统在不引起使用者特别注意的情况下，时刻采集用户的产生的数据，理想状态是所有的数据，包括（用户的行为，动作，言语，做的事，甚至用户的体温状态以及环境数据）。  能做好事情推荐或者更智能的决策的前提是得有足够准确，细致，完全的个人数据。在做好数据收集的情况下，再思考相应的算法或者方法，处理数据，得到合理的建议。  这次打算主要记录说话的语音数据，可能还会再加上心率，等其他辅助数据，得到一段时间的情感分析，并尽量输出相应的报告或图表供用户展示。  **1.2国内外研究现状**  1.2.1行业现状  自从1952年，AT&TBell实验室的Davis等人研制了第一个可十个英文数字的特定人语音增强系统一Audry系统1956年，美国普林斯顿大学RCA实验室的Olson和Belar等人研制出能10个单音节词的系统，该系统采用带通滤波器组获得的频谱参数作为语音增强特征。1959年，Fry和Denes等人尝试构建音素器来4个元音和9个辅音，并采用频谱分析和模式匹配进行决策。这就大大提高了语音识别的效率和准确度。从此计算机语音识别的受到了各国科研人员的重视并开始进入语音识别的研究。60年代，苏联的MaTIn等提出了语音结束点的端点检测，使语音识别水平明显上升;Vintsyuk提出了动态编程，这一提法在以后的识别中不可或缺。60年代末、70年代初的重要成果是提出了信号线性预测编码(LPC)技术和动态时间规整(DTW)技术，有效地解决了语音信号的特征提取和不等长语音匹配问题;同时提出了矢量量化(VQ)和隐马尔可夫模型(HMM)理论。语音识别技术与语音合成技术结合使人们能够摆脱键盘的束缚，取而代之的是以语音输入这样便于使用的、自然的、人性化的输入方式，它正逐步成为信息技术中人机接口的关键技术。  目前具有代表性的语音识别方法主要有动态时间规整技术(DTW)、隐马尔可夫模型(HMM)、矢量量化(VQ)、人工神经网络(ANN)、支持向量机(SVM)等方法。  动态时间规整算法(Dynamic TIme Warping，DTW)是在非特定人语音识别中一种简单有效的方法，该算法基于动态规划的思想，解决了发音长短不一的模板匹配问题，是语音识别技术中出现较早、较常用的一种算法。在应用DTW算法进行语音识别时，就是将已经预处理和分帧过的语音测试信号和参考语音模板进行比较以获取他们之间的相似度，按照某种距离测度得出两模板间的相似程度并选择最佳路径。  隐马尔可夫模型(HMM)是语音信号处理中的一种统计模型，是由Markov链演变来的，所以它是基于参数模型的统计识别方法。由于其模式库是通过反复训练形成的与训练输出信号吻合概率最大的最佳模型参数而不是预先储存好的模式样本，且其识别过程中运用待识别语音序列与HMM参数之间的似然概率达到最大值所对应的最佳状态序列作为识别输出，因此是较理想的语音识别模型。  矢量量化(Vector QuanTIzaTIon)是一种重要的信号压缩方法。与HMM相比，矢量量化主要适用于小词汇量、孤立词的语音识别中。其过程是将若干个语音信号波形或特征参数的标量数据组成一个矢量在多维空间进行整体量化。把矢量空间分成若干个小区域，每个小区域寻找一个代表矢量，量化时落入小区域的矢量就用这个代表矢量代替。矢量量化器的设计就是从大量信号样本中训练出好的码书，从实际效果出发寻找到好的失真测度定义公式，设计出最佳的矢量量化系统，用最少的搜索和计算失真的运算量实现最大可能的平均信噪比。  在实际的应用过程中，人们还研究了多种降低复杂度的方法，包括无记忆的矢量量化、有记忆的矢量量化和模糊矢量量化方法。  人工神经网络(ANN)是20世纪80年代末期提出的一种新的语音识别方法。其本质上是一个自适应非线性动力学系统，模拟了人类神经活动的原理，具有自适应性、并行性、鲁棒性、容错性和学习特性，其强大的分类能力和输入—输出映射能力在语音识别中都很有吸引力。其方法是模拟人脑思维机制的工程模型，它与HMM正好相反，其分类决策能力和对不确定信息的描述能力得到举世公认，但它对动态时间信号的描述能力尚不尽如人意，通常MLP分类器只能解决静态模式分类问题，并不涉及时间序列的处理。尽管学者们提出了许多含反馈的结构，但它们仍不足以刻画诸如语音信号这种时间序列的动态特性。由于ANN不能很好地描述语音信号的时间动态特性，所以常把ANN与传统识别方法结合，分别利用各自优点来进行语音识别而克服HMM和ANN各自的缺点。近年来结合神经网络和隐含马尔可夫模型的识别算法研究取得了显著进展，其识别率已经接近隐含马尔可夫模型的识别系统，进一步提高了语音识别的鲁棒性和准确率。  支持向量机(Support vector machine)是应用统计学理论的一种新的学习机模型，采用结构风险最小化原理(Structural Risk Minimization，SRM)，有效克服了传统经验风险最小化方法的缺点。兼顾训练误差和泛化能力，在解决小样本、非线性及高维模式识别方面有许多优越的性能，已经被广泛地应用到模式识别领域。  语音识别的研究工作可以追溯到20世纪50年代AT&T贝尔实验室的Audry系统，它是第一个可以识别十个英文数字的语音识别系统。  但真正取得实质性进展，并将其作为一个重要的课题开展研究则是在60年代末70年代初。这首先是因为计算机技术的发展为语音识别的实现提供了硬件和软件的可能，更重要的是语音信号线性预测编码(LPC)技术和动态时间规整(DTW)技术的提出，有效的解决了语音信号的特征提取和不等长匹配问题。这一时期的语音识别主要基于模板匹配原理，研究的领域局限在特定人，小词汇表的孤立词识别，实现了基于线性预测倒谱和DTW技术的特定人孤立词语音识别系统;同时提出了矢量量化(VQ)和隐马尔可夫模型(HMM)理论。  1.2.2 发展趋势  目前语音识别系统很难做到排除各种声学环境因素的影响，而人类语言在日常生活中的随意性和不确定性给语音识别系统造成极大的识别困难。所以，要应用现代技术智能化语音识别系统，以达到更好的识别效果；  目前语音识别系统使用的声学模型和语音模型过于局限，需要通过改进系统建模方法、提高搜索算法的效率来做到词汇量无限制和多重语言混合，减少词汇量对语音识别系统的限制；  语音识别系统在商业上的用途相当广泛，利用先进的微电子技术，将具有先进功能和性能的语音识别应用系统固化到更加微小的芯片或模块上，可以缩减成本，更方便的推广和使用。语音识别系统和微电子芯片技术的发展将引领信息技术革命到一个新的台阶。语音识别系统使人沟通更加自由，使人可以方便地享受到更多的社会信息资源和现代化服务。这必然会成为语音识别技术研究和应用的重要发展趋势。  **1.3研究意义**  语音识别系统在商业上的用途相当广泛，利用先进的微电子技术，将具有先进功能和性能的语音识别应用系统固化到更加微小的芯片或模块上，可以缩减成本，更方便的推广和使用。语音识别系统和微电子芯片技术的发展将引领信息技术革命到一个新的台阶。语音识别系统使人沟通更加自由，使人可以方便地享受到更多的社会信息资源和现代化服务。这必然会成为语音识别技术研究和应用的重要发展趋势。 | | | |
| 二、研究目标和内容 | | | |
| **2.1研究目标**  在不引起使用者特别注意的情况下，时刻采集用户的产生的数据，理想状态是所有的数据，包括（用户的行为，动作，言语，做的事，甚至用户的体温状态以及环境数据）。  能做好事情推荐或者更智能的决策的前提是得有足够准确，细致，完全的个人数据。在做好数据收集的情况下，再思考相应的算法或者方法，处理数据，得到合理的建议。  在数据收集方面能还原当天或者一个时间段的真实情况包括语言的大小和内容的准确性，用户能从客户端控制系统的开和关以及其他更多的设置（指定时间休眠等），通过报告或图表准确地得到自己的监控分析报告包括（情绪数据，说话的数据量大小）以及推荐的建议。  2**.2主要研究内容**  2.2.1背景调研与方案设计。  通过文献调研方法对语音处理管理系统的研究现状进行调研分析，了解行业发展现状，分析典型行业应用案例，研究本项目的理论意义与实践意义。并结合本项目设计目标提出相关功能需求，研究制定总体技术方案。  **2.2.2硬件系统设计。**  硬件电路是电路系统的重要组成部分，硬件电路设计是否合理直接影响电路系统的性能。硬件电路设计的一般分为设计需求分析、原理图设计、PCB设计、工艺文件处理等几个阶段，设计过程中的每一个细节都可能成为导致设计成功与失败的关键。根据总体技术方案设计相关硬件系统，包括多种语音感知模块、存储模块，无线传输模块以及显示模块等。通过Altium Designer等工具绘制硬件原理图，完成硬件模型搭建。  **2.2.3软件系统设计。**  从软件需求规格说明书出发，根据需求分析阶段确定的功能设计软件系统的整体结构、划分功能模块、确定每个模块的实现算法以及编写具体的代码，形成软件的具体设计方案，根据搭建硬件模型，结合总计研究目标设计相关软件功能模块。通过keil或者IAR开发平台编写硬件功能模块的控制代码，实现相应的信号采集、传输以及控制功能。设计后台控制APP以及相关数据库，实现家居环境的远程监控以及数据管理。  **2.2.4功能测试与优化。**  对设计的语音处理系统进行总体功能验证，并针对其中的不足之处提出优化方案。通过软件测试检验是否满足规定的需求和预期效果，主要采用白盒测试和黑盒测试来检查内部结构和具体实现的功能，采用静态测试和动态测试来测试程序执行的结果。 | | | |
| 三、研究方案 | | | |
| **3.1研究方法**  在这个语音分析系统中，主要包括感知层语音采集转存与发送、服务端数据的接收与分析，客户终端展示与设置 三个部分。接下来，我将分别对感知层语音采集转存与发送、服务端数据的接收与分析，客户终端展示与设置的实现进行阐述，提出大致的研究思路。系统整体结构图如图3-1所示。  Android  Internet  服务端  Stm32  Wi-Fi    4G  客户端 服务器端 感知层  图 3-1 系统结构图   1. 语音采集转存与发送处理：   利用VS1053芯片进行语音采集WAV格式文件，wav是最常用的数字化声音文件格式之一，其扩展名为“.wav”。符合RIFF(Resource Interchange File Format)文件规范，用于保存Windows平台的音频信息资源，被Windows平台及其应用程序所广泛支持。  然后将wav文件存入SD卡，再利用esp8266连入服务端，将数据源源不断的传给服务器。  服务端数据的接收与分析方案：  服务端数据的接收与分析是系统一个至关重要的环节，对整个系统的性能有着至关重要的影响。利用多线程监听单片机的连接，设置一套当前应用的专属传输层协议，方便数据的接收与传输，将接收到的wav格式进行语音识别（讯飞在线识别的相应API），识别出相应的文字，然后存入数据库或直接以文件形式存入服务器，调试相应的自然语言分析分析算法（关键字提取 情感分析 语义角色标注），得到当前情绪状况，当日说话量或更多信息，并且记录下来，发给客户终端展示。  客户终端展示与设置：  我将从系统客户端实现的角度出发，分别从客户端UI设计、客户端展示处理、控制系统的实现三个方面介绍预计的客户端的软件实现过程：  客户端UI设计  客户端的用户界而主要由登录界面、分析报告界面、个人偏好参数设置界面和工作状态控制界面。预计主体流程如图3-2所示。  客户端展示处理的实现  客户端展示处理是监控客户终端完成的主要工作之一，是移动客户端的主体部分。其主要完成分析好的数据的接收、并且生成相应的报告和图表展示给用户。  控制板块的实现  系统的控制部分主要体现在客户端对设备的控制。通过点击界面中相应按钮可以实现单片机的开始或结束录音操作，即当界面的控制按钮被容户端会发出控制指令，前端设备接收到控制指令后对控制指令进行解析，然后发出相应的控制信号以满足用户的需求。  登录界面  工作状态界面  个人偏好参数设置界面  分析报告图表  退出    点击登陆  点击 取消/保存    分析展示 点击设置      点击退出  图 3-2 客户端功能流程图 | | | |
| 三、进度计划（按月编制） | | | |
| **课题时间表：**  2021.01-2021.02：我将会加强对毕业设计所涉及技术进行着重了解，然后确定好整体的实现框架和具体的细节，特别是，数据采集，和自然语言分析的相关算法，需要系统的学习一段时间。另外，毕业设计所应用的Java语言和C语言也是我应该加强学习的部分。  本段时间我也将进行硬件设备采购，主要是选购通信芯片，主控芯片，音频采集芯片及其他配件。  2021.03-2021.04：整个系统的具体实现，软件开发及调试。  2021.04-2021.06：课题论文撰写，毕业设计成果展示，论文答辩。 | | | |
| 四、指导教师意见 | | | |
| □同意开题  □不同意开题   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | 指导教师签字： | |  | | | | |  | 年 |  | 月 |  | 日 | | | | |

备注：此报告应根据下达的毕业设计(论文)任务书，在指导教师的指导下由学生独立撰写，并于任务书下达后两周内完成。